

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :  C07K 14/705		A1	(11) International Publication Number: <b>WO 98/33818</b>  (43) International Publication Date: 6 August 1998 (06.08.98)
(21) International Application Number: PCT/US98/01422  (22) International Filing Date: 26 January 1998 (26.01.98)  (30) Priority Data: 08/792,014 31 January 1997 (31.01.97) US		(81) Designated States: AT, AU, BR, CA, CH, CN, DE, DK, ES, FI, GB, IL, JP, KR, MX, NO, NZ, RO, RU, SE, US, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 08/792,014 (CIP) Filed on 31 January 1997 (31.01.97)		Published <i>With international search report.</i>	
(71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94034 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). GOLI, Surya, K. [IN/US]; 3311 Cantamar Court, San Jose, CA 95135 (US).			
(74) Agent: BILLINGS, Lucy, J.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).			

(54) Title: HUMAN ANION CHANNEL

## (57) Abstract

The present invention provides a novel human anion channel (NANCH) and polynucleotides which identify and encode NANCH. The invention also provides genetically engineered expression vectors and host cells comprising the nucleic acid sequences encoding NANCH and a method for producing NANCH. The invention also provides for agonists, antibodies, or antagonists specifically binding NANCH, and their use, in the prevention and treatment of diseases associated with expression of NANCH. Additionally, the invention provides for the use of antisense molecules to polynucleotides encoding NANCH for the treatment of diseases associated with the expression of NANCH. The invention also provides diagnostic assays which utilize the polynucleotide, or fragments or the complement thereof, and antibodies specifically binding NANCH.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		

## HUMAN ANION CHANNEL

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of a novel human anion channel and to the use of these sequences in the diagnosis, prevention, and treatment of cancer 5 and developmental disorders.

### BACKGROUND ART

Chloride channels are found in the plasma membranes of virtually every cell in the body. Chloride channels mediate a variety of cellular functions including regulation of membrane potentials and absorption and secretion of ion across epithelial membranes. When present in 10 intracellular membranes of the Golgi apparatus and endocytic vesicles, chloride channels also regulate organelle pH (cf. Greger, R. (1988) *Annu. Rev. Physiol.* 50:111-122). Electrophysiological and pharmacological measurements including ion conductance, current-voltage relationships, and sensitivity to modulators suggest that different chloride channels exist 15 in muscles, neurons, fibroblasts, epithelial cells, and lymphocytes.

Several chloride channels have been cloned from mammalian tissues and cell lines. The sequences of these proteins are diverse and include a nicotinic acetylcholine receptor homolog, the cystic fibrosis transmembrane conductance regulator, and the highly acidic p64 protein with no significant homology to other proteins (Bernard, E.A. et al. (1987) *Trends Neurosci.* 16:502-509; Riordan, J.R. et al. (1989) *Science* 245:1066-1073; Landry, D. et al. (1993) *J. Biol. Chem.* 268:14948-14955). Many of the channels have sites for phosphorylation by one or more protein 20 kinases including protein kinase A, protein kinase C, tyrosine kinase, and casein kinase II which regulate chloride channel activity in epithelial cells.

The p64 protein was originally identified in bovine kidney cortex membranes by its affinity for indanyloxyacetic acid, a known inhibitor of epithelial chloride channels (Landry et al. 25 (1989) *Science* 244:1469-1472). Antibodies raised against the isolated p64 protein can deplete solubilized kidney membranes of all detectable chloride channel activity. Thus, p64 is likely to be a functional component of the kidney chloride channel (Redhead, R.C. et al. (1992) *Proc. Natl. Acad. Sci.* 89:3716-3720).

Northern blot analyses using the p64 clone as a probe detect related mRNAs, ranging in 30 size from ~2 kb to ~6.5 kb, in bovine kidney cortex, kidney medulla, liver, adrenal glands, brain, skeletal muscle, and heart. Most of these tissues have multiple transcripts that are capable of hybridizing to this probe. The diversity and relative abundance of these transcripts is tissue-

specific, and this suggests that the p64 transcripts are alternatively spliced and/or that a family of related genes are expressed (Landry et al., *supra*).

The sequence of p64 predicts an acidic, integral membrane protein which spans the membrane at least twice and has the amino terminus on the cytoplasmic side. The protein has 5 potential sites for phosphorylation by protein kinase C, tyrosine kinase, and casein kinase II, and a single site for N-linked glycosylation at Asp<sup>235</sup>.

The discovery of polynucleotides encoding a novel human anion channel, and the molecules themselves, provides a means to investigate the regulation of membrane potentials, intracellular pH, cell volume, signal transduction, and transepithelial ion transport in tissues 10 containing absorptive or secretory epithelia under normal and disease conditions. Discovery of a novel anion channel satisfies a need in the art by providing new compositions useful in diagnosing and treating cancer and developmental disorders.

#### DISCLOSURE OF THE INVENTION

The present invention features a novel human anion channel protein hereinafter 15 designated NANCH and characterized as having similarity to bovine p64 chloride channel and human P64CLCP.

Accordingly, the invention features a substantially purified NANCH having the amino acid sequence shown in SEQ ID NO:1.

One aspect of the invention features isolated and substantially purified polynucleotides 20 that encode NANCH. In a particular aspect, the polynucleotide is the nucleotide sequence of SEQ ID NO:2.

The invention also relates to a polynucleotide sequence comprising the complement of SEQ ID NO:2 or variants thereof. In addition, the invention features polynucleotide sequences which hybridize under stringent conditions to SEQ ID NO:2.

25 The invention additionally features nucleic acid sequences encoding polypeptides, oligonucleotides, peptide nucleic acids (PNA), fragments, portions or antisense molecules thereof, and expression vectors and host cells comprising polynucleotides that encode NANCH. The present invention also features antibodies which bind specifically to NANCH, and pharmaceutical compositions comprising substantially purified NANCH. The invention also 30 features the use of agonists and antagonists of NANCH. The invention also features methods for treating disorders which are associated with NANCH, and for detecting a polynucleotide which encodes NANCH.

**BRIEF DESCRIPTION OF DRAWINGS**

Figures 1A, 1B and 1C show the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:2) of NANCH. The alignment was produced using MacDNASIS PRO™ software (Hitachi Software Engineering Co., Ltd., San Bruno, CA).

5 Figures 2A, 2B and 2C show the amino acid sequence alignments among NANCH (SEQ ID NO:1), bovine p64 chloride channel (GI 289404; SEQ ID NO:3), and human P64CLCP (GI 895845; SEQ ID NO:4). The alignment was produced using the multisequence alignment program of DNASTAR™ software (DNASTAR Inc, Madison WI).

Figures 3A and 3B show the hydrophobicity plots (produced using the protein analysis 10 program of DNASTAR software) for NANCH, SEQ ID NO:1, and bovine p64 chloride channel, SEQ ID NO:3, respectively; the positive X axis reflects amino acid position, and the negative Y axis, hydrophobicity.

**MODES FOR CARRYING OUT THE INVENTION**

Before the present proteins, nucleotide sequences, and methods are described, it is 15 understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

20 It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" includes a plurality of such host cells, reference to the "antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

25 Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by 30 reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate

such disclosure by virtue of prior invention.

#### DEFINITIONS

“Nucleic acid sequence” as used herein refers to an oligonucleotide, nucleotide, or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand.

Similarly, “amino acid sequence” as used herein refers to an oligopeptide, peptide, polypeptide, or protein sequence, and fragments or portions thereof, and to naturally occurring or synthetic molecules.

Where “amino acid sequence” is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, “amino acid sequence” and like terms, such as “polypeptide” or “protein” are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

“Peptide nucleic acid”, as used herein, refers to a molecule which comprises an oligomer to which an amino acid residue, such as lysine, and an amino group have been added. These small molecules, also designated anti-gene agents, stop transcript elongation by binding to their complementary strand of nucleic acid (Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63).

NANCH, as used herein, refers to the amino acid sequences of substantially purified NANCH obtained from any species, particularly mammalian, including bovine, ovine, porcine, murine, equine, and preferably human, from any source whether natural, synthetic, semi-synthetic, or recombinant.

“Consensus”, as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, or which has been extended using XL-PCR™ (Perkin Elmer, Norwalk, CT) in the 5' and/or the 3' direction and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte clone using the GELVIEW™ Fragment Assembly system (GCG, Madison, WI), or which has been both extended and assembled.

A “variant” of NANCH, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have “conservative” changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant may have “nonconservative” changes, e.g., replacement of a glycine with a tryptophan. Similar minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted,

inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

A "deletion", as used herein, refers to a change in either amino acid or nucleotide sequence in which one or more amino acid or nucleotide residues, respectively, are absent.

5 An "insertion" or "addition", as used herein, refers to a change in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid or nucleotide residues, respectively, as compared to the naturally occurring molecule.

A "substitution", as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

10 The term "biologically active", as used herein, refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic NANCH, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

15 The term "agonist", as used herein, refers to a molecule which, when bound to NANCH, causes a change in NANCH which modulates the activity of NANCH. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to NANCH.

The terms "antagonist" or "inhibitor", as used herein, refer to a molecule which, when bound to NANCH, blocks or modulates the biological or immunological activity of NANCH.

20 Antagonists and inhibitors may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to NANCH.

The term "modulate", as used herein, refers to a change or an alteration in the biological activity of NANCH. Modulation may be an increase or a decrease in protein activity, a change in binding characteristics, or any other change in the biological, functional or immunological

25 properties of NANCH.

The term "mimetic", as used herein, refers to a molecule, the structure of which is developed from knowledge of the structure of NANCH or portions thereof and, as such, is able to effect some or all of the actions of bovine p64-like molecules.

The term "derivative", as used herein, refers to the chemical modification of a nucleic acid encoding NANCH or the encoded NANCH. Illustrative of such modifications would be replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative would encode a polypeptide which retains essential biological characteristics of the natural molecule.

The term "substantially purified", as used herein, refers to nucleic or amino acid sequences that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated.

5 "Amplification" as used herein refers to the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction (PCR) technologies well known in the art (Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY).

10 The term "hybridization", as used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

15 The term "hybridization complex", as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary G and C bases and between complementary A and T bases; these hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in solution (e.g.,  $C_{0,t}$  or  $R_{0,t}$  analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., membranes, filters, chips, pins or glass slides to which cells have been fixed for *in situ* hybridization).

20 The terms "complementary" or "complementarity", as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For example, for the sequence "A-G-T" binds to the complementary sequence "T-C-A". Complementarity between two single-stranded molecules may be "partial", in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has 25 significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands.

20 The term "homology", as used herein, refers to a degree of complementarity. There may be partial homology or complete homology (i.e., identity). A partially complementary sequence is one that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid; it is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be

examined using a hybridization assay (Southern or northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous sequence or probe to the target sequence under conditions of low stringency. This is not to say that 5 conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% identity); in the absence of non-specific binding, the probe will not hybridize to the second non- 10 complementary target sequence.

As known in the art, numerous equivalent conditions may be employed to comprise either low or high stringency conditions. Factors such as the length and nature (DNA, RNA, base composition) of the sequence, nature of the target (DNA, RNA, base composition, presence in solution or immobilization, etc.), and the concentration of the salts and other components (e.g., 15 the presence or absence of formamide, dextran sulfate and/or polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of either low or high stringency different from, but equivalent to, the above listed conditions.

The term "stringent conditions", as used herein, is the "stringency" which occurs within a range from about  $T_m - 5^\circ\text{C}$  ( $5^\circ\text{C}$  below the melting temperature ( $T_m$ ) of the probe) to about  $20^\circ\text{C}$  20 to  $25^\circ\text{C}$  below  $T_m$ . As will be understood by those of skill in the art, the stringency of hybridization may be altered in order to identify or detect identical or related polynucleotide sequences.

The term "antisense", as used herein, refers to nucleotide sequences which are complementary to a specific DNA or RNA sequence. The term "antisense strand" is used in 25 reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules may be produced by any method, including synthesis by ligating the gene(s) of interest in a reverse orientation to a viral promoter which permits the synthesis of a complementary strand. Once introduced into a cell, this transcribed strand combines with natural sequences produced by the cell to form duplexes. These duplexes then block either the further transcription 30 or translation. In this manner, mutant phenotypes may be generated. The designation "negative" is sometimes used in reference to the antisense strand, and "positive" is sometimes used in reference to the sense strand.

The term "portion", as used herein, with regard to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino acid sequence minus one amino acid. Thus, a protein "comprising at least a portion of the amino acid sequence of SEQ ID NO:1" encompasses the 5 full-length human NANCH and fragments thereof.

"Transformation", as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. It may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is 10 selected based on the host cell being transformed and may include, but is not limited to, viral infection, electroporation, lipofection, and particle bombardment. Such "transformed" cells include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome. They also include cells which transiently express the inserted DNA or RNA for limited periods of time.

15 The term "antigenic determinant", as used herein, refers to that portion of a molecule that makes contact with a particular antibody (i.e., an epitope). When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as antigenic determinants. An antigenic 20 determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The terms "specific binding" or "specifically binding", as used herein, in reference to the interaction of an antibody and a protein or peptide, mean that the interaction is dependent upon the presence of a particular structure (i.e., the antigenic determinant or epitope) on the protein; in 25 other words, the antibody is recognizing and binding to a specific protein structure rather than to proteins in general. For example, if an antibody is specific for epitope "A", the presence of a protein containing epitope A (or free, unlabeled A) in a reaction containing labeled "A" and the antibody will reduce the amount of labeled A bound to the antibody.

The term "sample", as used herein, is used in its broadest sense. A biological sample 30 suspected of containing nucleic acid encoding NANCH or fragments thereof may comprise a cell, chromosomes isolated from a cell (e.g., a spread of metaphase chromosomes), genomic DNA (in solution or bound to a solid support such as for Southern analysis), RNA (in solution or bound to

a solid support such as for northern analysis), cDNA (in solution or bound to a solid support), an extract from cells or a tissue, and the like.

The term "correlates with expression of a polynucleotide", as used herein, indicates that the detection of the presence of ribonucleic acid that is similar to SEQ ID NO:2 by northern analysis is indicative of the presence of mRNA encoding NANCH in a sample and thereby correlates with expression of the transcript from the polynucleotide encoding the protein.

"Alterations" in the polynucleotide of SEQ ID NO: 2, as used herein, comprise any alteration in the sequence of polynucleotides encoding NANCH including deletions, insertions, and point mutations that may be detected using hybridization assays. Included within this definition is the detection of alterations to the genomic DNA sequence which encodes NANCH (e.g., by alterations in the pattern of restriction fragment length polymorphisms capable of hybridizing to SEQ ID NO:2), the inability of a selected fragment of SEQ ID NO:2 to hybridize to a sample of genomic DNA (e.g., using allele-specific oligonucleotide probes), and improper or unexpected hybridization, such as hybridization to a locus other than the normal chromosomal locus for the polynucleotide sequence encoding NANCH (e.g., using fluorescent *in situ* hybridization (FISH) to metaphase chromosomes spreads).

As used herein, the term "antibody" refers to intact molecules as well as fragments thereof, such as Fa, F(ab')<sub>2</sub>, and Fv, which are capable of binding the epitopic determinant. Antibodies that bind NANCH polypeptides can be prepared using intact polypeptides or fragments containing small peptides of interest as the immunizing antigen. The polypeptide or peptide used to immunize an animal can be derived from the transition of RNA or synthesized chemically, and can be conjugated to a carrier protein, if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin and thyroglobulin. The coupled peptide is then used to immunize the animal (e.g., a mouse, a rat, or a rabbit).

The term "humanized antibody", as used herein, refers to antibody molecules in which amino acids have been replaced in the non-antigen binding regions in order to more closely resemble a human antibody, while still retaining the original binding ability.

#### THE INVENTION

The invention is based on the discovery of a novel human anion channel (NANCH), the polynucleotides encoding NANCH, and the use of these compositions for the diagnosis, prevention, or treatment of cancer and developmental disorders.

The nucleic acid sequence encoding the human NANCH of the present invention was first

identified in Incyte Clone 259592 from an hNT2 neuronal cell line cDNA library (HNT2RAT01) through a computer-generated search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 040177 (TBLYN01); 259592 (HNT2RAT01); 311528 5 (LUNGNOT02); 760901 (BRAITUT02); 1522436 (BLADTUT04); and 2134172 (ENDCNOT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1, as shown in Figures 1A, 1B and 1C. NANCH is 253 amino acids in length and has potential transmembrane domains at residues 19-54 (which is long enough to span 10 the membrane twice) and residues 187-201, and a potential  $\beta$ -sheet-forming transmembrane domain at residues 116-129. NANCH contains consensus sequences for phosphorylation by protein kinase C and casein kinase II at S<sup>236</sup>, tyrosine kinase at Y<sup>244</sup>, and cAMP-dependent protein kinase at T<sup>252</sup>. These potential phosphorylation sites are all located near the C-terminus of 15 NANCH. NANCH has chemical and structural homology with bovine p64 chloride channel (GI 289404; SEQ ID NO:3) and human P64CLCP (GI 895845; SEQ ID NO:4). This suggests that NANCH is involved in the regulation of transepithelial ion transport in tissues containing absorptive or secretory epithelia.

In the region of homology, NANCH shares 76% and 65% identity with GI 289404 and GI 895845, respectively (Figures 2A, 2B and 2C). GI 289404 is considerably longer than NANCH 20 or GI 895845, and its first 200 amino acid residues have no significant counterparts in either NANCH or GI 895845. This may reflect species differences, alternative splicing, and/or transcription from different members of the gene family. As illustrated by Figures 3A and 3B, NANCH and GI 289404 have similar hydrophobicity plots in the region of homology. Northern analysis reveals the expression of the NANCH sequence in libraries derived from vascular 25 endothelial cells and left ventricle of the heart, and in tissues involved in secretion and adsorption including stomach, lung, kidney, tongue, colon, large intestine, pancreas, adrenal gland, thyroid, bladder, and pancreas. Sixteen of the libraries (55%) are derived from tumors and fetal tissues, suggesting that NANCH is associated with regulation of cell growth.

The invention also encompasses NANCH variants. A preferred NANCH variant is one 30 having at least 80%, and more preferably 90%, amino acid sequence identity to the NANCH amino acid sequence (SEQ ID NO:1). A most preferred NANCH variant is one having at least 95% amino acid sequence identity to SEQ ID NO:1.

The invention also encompasses polynucleotides which encode NANCH. Accordingly, any nucleic acid sequence which encodes the amino acid sequence of NANCH can be used to generate recombinant molecules which express NANCH. In a particular embodiment, the invention encompasses the polynucleotide comprising the nucleic acid sequence of SEQ ID NO:2 as shown in Figures 1A, 1B and 1C.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding NANCH, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring NANCH, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode NANCH and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring NANCH under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding NANCH or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding NANCH and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences, or portions thereof, which encode NANCH and its derivatives, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art at the time of the filing of this application. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding NANCH or any portion thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequences, and in particular, those shown in SEQ ID NO:2, under various conditions of stringency. Hybridization conditions are based on the melting

temperature (T<sub>m</sub>) of the nucleic acid binding complex or probe, as taught in Wahl, G.M. and S.L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A.R. (1987; Methods Enzymol. 152:507-511), and may be used at a defined stringency.

Altered nucleic acid sequences encoding NANCH which are encompassed by the invention include deletions, insertions, or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent NANCH. The encoded protein may also contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent NANCH. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, 5 hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological activity of NANCH is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged amino acids may include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; 10 serine and threonine; phenylalanine and tyrosine.

15

Also included within the scope of the present invention are alleles of the genes encoding NANCH. As used herein, an "allele" or "allelic sequence" is an alternative form of the gene which may result from at least one mutation in the nucleic acid sequence. Alleles may result in altered mRNAs or polypeptides whose structure or function may or may not be altered. Any 20 given gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

Methods for DNA sequencing which are well known and generally available in the art 25 may be used to practice any embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of recombinant polymerases and proofreading exonucleases such as the ELONGASE Amplification System marketed by Gibco BRL (Gaithersburg, MD).

30 Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA) and the ABI 377 DNA sequencers (Perkin Elmer).

The nucleic acid sequences encoding NANCH may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed, "restriction-site" PCR, uses universal primers to retrieve unknown sequence adjacent 5 to a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). In particular, genomic DNA is first amplified in the presence of primer to linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse 10 transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). The primers may be designed using OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 15 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA 20 (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may also be used to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before performing PCR.

Another method which may be used to retrieve unknown sequences is that of Parker, J.D. 25 et al. (1991; Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk in genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they 30 will contain more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into the 5' and 3'

non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, 5 four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled device camera. Output/light intensity may be converted to electrical signal using appropriate software (e.g. Genotyper<sup>TM</sup> and Sequence Navigator<sup>TM</sup>, Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially 10 preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode NANCH, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct expression of NANCH in appropriate host cells. Due to 15 the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express NANCH.

As will be understood by those of skill in the art, it may be advantageous to produce NANCH-encoding nucleotide sequences possessing non-naturally occurring codons. For 20 example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods 25 generally known in the art in order to alter NANCH encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation 30 patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding NANCH may be ligated to a heterologous sequence to encode a fusion

protein. For example, to screen peptide libraries for inhibitors of NANCH activity, it may be useful to encode a chimeric NANCH protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the NANCH encoding sequence and the heterologous protein sequence, so that NANCH may be 5 cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding NANCH may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M.H. et al. (1980) *Nuc. Acids Res. Symp. Ser.* 215-223, Horn, T. et al. (1980) *Nuc. Acids Res. Symp. Ser.* 225-232).

Alternatively, the protein itself may be produced using chemical methods to synthesize the amino 10 acid sequence of NANCH, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J.Y. et al. (1995) *Science* 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer).

The newly synthesized peptide may be substantially purified by preparative high 15 performance liquid chromatography (e.g., Creighton, T. (1983) *Proteins, Structures and Molecular Principles*, W.H. Freeman and Co., New York, NY). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, T., *supra*). Additionally, the amino acid sequence of NANCH, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with 20 sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active NANCH, the nucleotide sequences encoding NANCH or functional equivalents, may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

25 Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding NANCH and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, 30 Plainview, NY, and Ausubel, F.M. et al. (1989) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY.

A variety of expression vector/host systems may be utilized to contain and express

sequences encoding NANCH. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

The "control elements" or "regulatory sequences" are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, La Jolla, CA) or pSport1™ plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding NANCH, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for NANCH. For example, when large quantities of NANCH are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional E. coli cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding NANCH may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of  $\beta$ -galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors (Promega, Madison, WI) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to

include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, 5 see Ausubel et al. (supra) and Grant et al. (1987) *Methods Enzymol.* 153:516-544.

In cases where plant expression vectors are used, the expression of sequences encoding NANCH may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant 10 promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, 15 Hobbs, S. or Murry, L.E. in *McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York, NY; pp. 191-196.

An insect system may also be used to express NANCH. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding NANCH 20 may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of NANCH will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which NANCH may be expressed (Engelhard, E.K. et al. (1994) *Proc. Nat. Acad. Sci.* 25 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding NANCH may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may 30 be used to obtain a viable virus which is capable of expressing NANCH in infected host cells (Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression

in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding NANCH. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding NANCH, its initiation codon, and upstream 5 sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation 10 codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of 15 the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, HEK293, and WI38, which 20 have specific cellular machinery and characteristic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express NANCH may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression 25 elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using 30 tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al.

(1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1980) Cell 22:817-23) genes which can be employed in tk<sup>-</sup> or aprt<sup>-</sup> cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); 5 npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman, S.C. and R.C. Mulligan 10 (1988) Proc. Natl. Acad. Sci. 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins,  $\beta$ -glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131).

15 Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding NANCH is inserted within a marker gene sequence, recombinant cells containing sequences encoding NANCH can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding 20 NANCH under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

25 Alternatively, host cells which contain the nucleic acid sequence encoding NANCH and express NANCH may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

The presence of polynucleotide sequences encoding NANCH can be detected by 30 DNA-DNA or DNA-RNA hybridization or amplification using probes or portions or fragments of polynucleotides encoding NANCH. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding NANCH to detect transformants containing DNA or RNA encoding NANCH. As used herein "oligonucleotides" or "oligomers" refer to a nucleic acid sequence of at least about 10 nucleotides and as many as about

60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides, which can be used as a probe or amplifier.

A variety of protocols for detecting and measuring the expression of NANCH, using either polyclonal or monoclonal antibodies specific for the protein are known in the art.

5 Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on NANCH is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St

10 Paul, MN) and Maddox, D.E. et al. (1983; *J. Exp. Med.* 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding NANCH include oligolabeling, nick translation, end-labeling or PCR amplification using a 15 labeled nucleotide. Alternatively, the sequences encoding NANCH, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, 20 Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp., Cleveland, OH). Suitable reporter molecules or labels, which may be used, include radionuclides, enzymes, 25 fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding NANCH may be cultured 25 under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode NANCH may be designed to contain signal sequences which direct secretion of NANCH through a prokaryotic or eukaryotic cell membrane. 30 Other recombinant constructions may be used to join sequences encoding NANCH to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides

such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San 5 Diego, CA) between the purification domain and NANCH may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing NANCH and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography as described in Porath, J. et al. (1992, *Prot. Exp. Purif.* 3: 263-281) while the 10 enterokinase cleavage site provides a means for purifying NANCH from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. et al. (1993; *DNA Cell Biol.* 12:441-453).

In addition to recombinant production, fragments of NANCH may be produced by direct peptide synthesis using solid-phase techniques Merrifield J. (1963) *J. Am. Chem. Soc.* 15 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of NANCH may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

#### **THERAPEUTICS**

20 Chemical and structural homology exists among NANCH, bovine p64, and human P64CLCP. This homology together with its northern analysis suggests that NANCH has a role in development. Accordingly, lack of NANCH may be implicated in developmental disorders.

Therefore, in one embodiment, NANCH or a fragment or derivative thereof may be administered to a subject to treat developmental disorders. Such disorders may include, but are 25 not limited to, incomplete development of any organ, including the heart, lungs, kidneys or liver in neonates, as well as congenital defects.

In another embodiment, a vector capable of expressing NANCH, or a fragment or a derivative thereof, may also be administered to a subject to treat the developmental disorders listed above.

30 NANCH also appears to have a role in the development of cancer. Therefore, in one embodiment, antagonists or inhibitors of NANCH may be administered to a subject to treat or prevent any type of cancer including, but not limited to, those of the heart, stomach, lung, skin,

kidney, tongue, colon, large intestine, liver, pancreas, adrenal gland, thyroid, bladder, and pancreas. In one aspect, antibodies which are specific for NANCH may be used directly as an antagonist, or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express NANCH.

5 In another embodiment, a vector expressing antisense of the polynucleotide encoding NANCH may be administered to a subject to treat or prevent the types of cancer including, but not limited to, those listed above.

In other embodiments, any of the therapeutic proteins, antagonists, antibodies, agonists, antisense sequences or vectors described above may be administered in combination with other 10 appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing 15 the potential for adverse side effects.

Antagonists or inhibitors of NANCH may be produced using methods which are generally known in the art. In particular, purified NANCH may be used to produce antibodies or to screen 20 libraries of pharmaceutical agents to identify those which specifically bind NANCH. Antibodies to NANCH may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with NANCH or any fragment or 25 oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli 30 Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the peptides, fragments, or oligopeptides used to induce antibodies to NANCH have an amino acid sequence consisting of at least five amino acids, and more

preferably at least 10 amino acids. It is also preferable that they are identical to a portion of the amino acid sequence of the natural protein, and they may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of NANCH amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against 5 the chimeric molecule.

Monoclonal antibodies to NANCH may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985)

10 *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci.* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; Takeda, S. et al. (1985) *Nature* 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce NANCH-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton 15 D.R. (1991) *Proc. Natl. Acad. Sci.* 88:11120-3).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci.* 86: 3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299).

25 Antibody fragments which contain specific binding sites for NANCH may also be generated. For example, such fragments include, but are not limited to, the F(ab')2 fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal 30 Fab fragments with the desired specificity (Huse, W.D. et al. (1989) *Science* 254:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays

using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between NANCH and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering NANCH epitopes is preferred, but a 5 competitive binding assay may also be employed (Maddox, *supra*).

In another embodiment of the invention, the polynucleotides encoding NANCH, or any fragment thereof, or antisense molecules, may be used for therapeutic purposes. In one aspect, antisense to the polynucleotide encoding NANCH may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with 10 sequences complementary to polynucleotides encoding NANCH. Thus, antisense molecules may be used to modulate NANCH activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligomers or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding NANCH.

15 Expression vectors derived from retro viruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct recombinant vectors which will express antisense molecules complementary to the polynucleotides of the gene encoding NANCH. These techniques are described both in 20 Sambrook et al. (*supra*) and in Ausubel et al. (*supra*).

Genes encoding NANCH can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide or fragment thereof which encodes NANCH. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may 25 continue to transcribe RNA molecules until they are disabled by endogenous nucleases.

Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA, or PNA, to the control regions of the gene encoding NANCH, 30 i.e., the promoters, enhancers, and introns. Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is

useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. 5 Kisco, NY). The antisense molecules may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. 10 Examples which may be used include engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding NANCH.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following 15 sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

20 Antisense molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding NANCH. Such DNA sequences may be incorporated into a wide 25 variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize antisense RNA constitutively or inducibly can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' 30 ends of the molecule or the use of phosphorothioate or 2' O-methyl phosphodiester linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine,

queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally 5 suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection and by liposome injections may be achieved using methods which are well known in the art.

Any of the therapeutic methods described above may be applied to any subject in need of 10 such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of 15 NANCH, antibodies to NANCH, mimetics, agonists, antagonists, or inhibitors of NANCH. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs 20 or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

25 In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, 30 PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral

administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active 5 compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums 10 including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, 15 polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. 20 Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in 25 aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles 30 include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly

concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a 5 manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the 10 corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of 15 NANCH, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

20 For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

25 A therapeutically effective dose refers to that amount of active ingredient, for example NANCH or fragments thereof, antibodies of NANCH, agonists, antagonists or inhibitors of NANCH, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose 30 lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell

culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

5 The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction  
10 sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

15 Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

#### DIAGNOSTICS

20 In another embodiment, antibodies which specifically bind NANCH may be used for the diagnosis of conditions or diseases characterized by expression of NANCH, or in assays to monitor patients being treated with NANCH, agonists, antagonists or inhibitors. The antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for NANCH include methods which utilize the antibody and a  
25 label to detect NANCH in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules which are known in the art may be used, several of which are described above.

30 A variety of protocols including ELISA, RIA, and FACS for measuring NANCH are known in the art and provide a basis for diagnosing altered or abnormal levels of NANCH expression. Normal or standard values for NANCH expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with

antibody to NANCH under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, but preferably by photometric, means. Quantities of NANCH expressed in subject, control and disease, samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values 5 establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding NANCH may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, antisense RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of NANCH may be 10 correlated with disease. The diagnostic assay may be used to distinguish between absence, presence, and excess expression of NANCH, and to monitor regulation of NANCH levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding NANCH or closely related 15 molecules, may be used to identify nucleic acid sequences which encode NANCH. The specificity of the probe, whether it is made from a highly specific region, e.g., 10 unique nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences encoding 20 NANCH, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the NANCH encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence of SEQ ID NO:2 or from genomic sequence including promoter, enhancer 25 elements, and introns of the naturally occurring NANCH.

Means for producing specific hybridization probes for DNAs encoding NANCH include the cloning of nucleic acid sequences encoding NANCH or NANCH derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA 30 polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as 32P or 35S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the

like.

Polynucleotide sequences encoding NANCH may be used for the diagnosis of conditions or diseases which are associated with expression of NANCH. Examples of such conditions or diseases include cancers of the heart, stomach, lung, skin, kidney, tongue, colon, large intestine, 5 liver, pancreas, adrenal gland, thyroid, bladder, and pancreas. The polynucleotide sequences encoding NANCH may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dip stick, pin, ELISA or chip assays utilizing fluids or tissues from patient biopsies to detect altered NANCH expression. Such qualitative or quantitative methods are well known in the art.

10 In a particular aspect, the nucleotide sequences encoding NANCH may be useful in assays that detect activation or induction of various cancers, particularly those mentioned above. The nucleotide sequences encoding NANCH may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated 15 and compared with a standard value. If the amount of signal in the biopsied or extracted sample is significantly altered from that of a comparable control sample, the nucleotide sequences have hybridized with nucleotide sequences in the sample, and the presence of altered levels of nucleotide sequences encoding NANCH in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic 20 treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of an individual patient.

In order to provide a basis for the diagnosis of disease associated with expression of NANCH, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with 25 a sequence, or a fragment thereof, which encodes NANCH, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for 30 disease. Deviation between standard and subject values is used to establish the presence of disease.

Once disease is established and a treatment protocol is initiated, hybridization assays may

be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

5 With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or

10 further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding NANCH may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced from a recombinant source. Oligomers will preferably consist of two nucleotide sequences, one with sense orientation (5'>3') and another with antisense (3'<5'),

15 employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of NANCH include

20 radiolabeling or biotinylation of nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby, P.C. et al. (1993) J. Immunol. Methods, 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or

25 colorimetric response gives rapid quantitation.

In another embodiment of the invention, the nucleic acid sequences which encode NANCH may also be used to generate hybridization probes which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. Such

30 techniques include FISH, FACS, or artificial chromosome constructions, such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price, C.M. (1993) Blood Rev. 7:127-134, and Trask, B.J. (1991)

Trends Genet. 7:149-154.

FISH (as described in Verma et al. (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York, NY) may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be 5 found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of the gene encoding NANCH on a physical chromosomal map and a specific disease, or predisposition to a specific disease, may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier, or affected individuals.

10        In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts 15 thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti, R.A. et al. (1988) *Nature* 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide 20 sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

      In another embodiment of the invention, NANCH, its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, 25 affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between NANCH and the agent being tested, may be measured.

      Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, as applied to NANCH large numbers 30 of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with NANCH, or fragments thereof, and washed. Bound NANCH is then detected by methods well known in the art. Purified NANCH

can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which 5 neutralizing antibodies capable of binding NANCH specifically compete with a test compound for binding NANCH. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with NANCH.

In additional embodiments, the nucleotide sequences which encode NANCH may be used in any molecular biology techniques that have yet to be developed, provided the new techniques 10 rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

#### INDUSTRIAL APPLICABILITY

##### 15 I HNT2RAT01 cDNA Library Construction

The hNT2 cell line exhibits characteristics of a committed neuronal precursor cell which is still at an early stage of development. The hNT2 cell line can be induced by retinoic acid (RA) to differentiate, as described in Andrews, P.W. (1984; Dev. Biol. 103:285-293). For purposes of this invention, hNT2 cells were induced with RA by suspending the cells in Dulbecco's modified 20 Eagle's medium (DMEM) including 10% fetal bovine serum plus penicillin/ streptomycin and treating with 10  $\mu$ M RA for 24 hours. The cells were harvested immediately thereafter.

The HNT2RAT01 cDNA library prepared from this treated cell line was prepared by Stratagene (Cat. No. 937231). The cDNA library was constructed by essentially the following procedure. cDNAs were primed using oligo d(T) and size fractionated to isolate fragments of 25 500 bp and larger. Synthetic adapter oligonucleotides were ligated onto the cDNA molecules enabling them to be inserted into the Uni-ZAP<sup>TM</sup> vector system (Stratagene). The quality of the cDNA library was screened using DNA probes, and the pBluescript phagemid (Stratagene) was then excised. Subsequently, the custom-constructed library phage particles were infected into E. coli host strain XL1-Blue<sup>®</sup> (Stratagene).

##### 30 II Isolation and Sequencing of cDNA Clones

Plasmid or phagemid DNA was released from cells and purified using the Miniprep Kit (Cat. No. 77468; Advanced Genetic Technologies Corporation, Gaithersburg MD). This kit

consists of a 96 well block with reagents for 960 purifications. The recommended protocol was employed except for the following changes: 1) the 96 wells were each filled with only 1 ml of sterile Terrific Broth (Cat. No. 22711, LIFE TECHNOLOGIES™, Gaithersburg MD) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) the bacteria were cultured for 24 hours after the 5 wells were inoculated and then lysed with 60  $\mu$ l of lysis buffer; 3) a centrifugation step employing the Beckman GS-6R at 2900 rpm for 5 min was performed before the contents of the block were added to the primary filter plate; and 4) the optional step of adding isopropanol to TRIS buffer was not routinely performed. After the last step in the protocol, samples were transferred to a Beckman 96-well block for storage.

10 Alternative methods of purifying plasmid DNA include the use of MAGIC MINIPREPSTM DNA Purification System (Cat. No. A7100, Promega) or QIAwell™-8 Plasmid, QIAwell PLUS DNA and QIAwell ULTRA DNA Purification Systems (Qiagen, Inc.).

The cDNAs were sequenced by the method of Sanger F. and A.R. Coulson (1975; J. Mol. Biol. 94:441f), using a Hamilton Micro Lab 2200 (Hamilton, Reno NV) in combination with four 15 Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown MA) and Applied Biosystems 377 or 373 DNA Sequencing Systems (Perkin Elmer) and reading frame was determined.

### III Homology Searching of cDNA Clones and Their Deduced Proteins

Each cDNA was compared to sequences in GenBank using a search algorithm developed by Applied Biosystems and incorporated into the INHERIT™ 670 sequence analysis system. In 20 this algorithm, Pattern Specification Language (TRW Inc, Los Angeles, CA) was used to determine regions of homology. The three parameters that determine how the sequence comparisons run were window size, window offset, and error tolerance. Using a combination of these three parameters, the DNA database was searched for sequences containing regions of homology to the query sequence, and the appropriate sequences were scored with an initial value. 25 Subsequently, these homologous regions were examined using dot matrix homology plots to distinguish regions of homology from chance matches. Smith-Waterman alignments were used to display the results of the homology search.

Peptide and protein sequence homologies were ascertained using the INHERIT- 670 sequence analysis system using the methods similar to those used in DNA sequence homologies. 30 Pattern Specification Language and parameter windows were used to search protein databases for sequences containing regions of homology which were scored with an initial value. Dot-matrix homology plots were examined to distinguish regions of significant homology from chance

matches.

BLAST, which stands for Basic Local Alignment Search Tool (Altschul, S.F. (1993) J. Mol. Evol. 36:290-300; Altschul et al. (1990) J. Mol. Biol. 215:403-410), was used to search for local sequence alignments. BLAST produces alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying homologs. BLAST is useful for matches which do not contain gaps. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user. The BLAST approach is to look for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output.

#### IV Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook et al., *supra*).

Analogous computer techniques using BLAST (Altschul, S.F. 1993 and 1990, *supra*) are used to search for identical or related molecules in nucleotide databases such as GenBank or the LIFESEQ™ database (Incyte Pharmaceuticals). This analysis is much faster than multiple, membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

30

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous molecules are usually

identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analysis are reported as a list of libraries in which the transcript encoding NANCH occurs. Abundance and percent abundance are also reported. Abundance 5 directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

#### V Extension of NANCH-Encoding Polynucleotides to Full Length or to Recover Regulatory Sequences

10 Full length NANCH-encoding nucleic acid sequence (SEQ ID NO:2) is used to design oligonucleotide primers for extending a partial nucleotide sequence to full length or for obtaining 5' or 3', intron or other control sequences from genomic libraries. One primer is synthesized to initiate extension in the antisense direction (XLR) and the other is synthesized to extend sequence in the sense direction (XLF). Primers are used to facilitate the extension of the known sequence 15 "outward" generating amplicons containing new, unknown nucleotide sequence for the region of interest. The initial primers are designed from the cDNA using OLIGO 4.06 (National Biosciences), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer 20 dimerizations is avoided.

The original, selected cDNA libraries, or a human genomic library are used to extend the sequence; the latter is most useful to obtain 5' upstream regions. If more extension is necessary or desired, additional sets of primers are designed to further extend the known region.

By following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing 25 the enzyme and reaction mix, high fidelity amplification is obtained. Beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, PCR is performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA) and the following parameters:

30	Step 1	94°C for 1 min (initial denaturation)
	Step 2	65°C for 1 min
	Step 3	68°C for 6 min
	Step 4	94°C for 15 sec
	Step 5	65°C for 1 min
	Step 6	68°C for 7 min

Step 7	Repeat step 4-6 for 15 additional cycles
Step 8	94°C for 15 sec
Step 9	65°C for 1 min
Step 10	68°C for 7:15 min
5 Step 11	Repeat step 8-10 for 12 cycles
Step 12	72°C for 8 min
Step 13	4°C (and holding)

A 5-10  $\mu$ l aliquot of the reaction mixture is analyzed by electrophoresis on a low 10 concentration (about 0.6-0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products are selected and removed from the gel. Further purification involves using a commercial gel extraction method such as QIAQuick™ (QIAGEN Inc., Chatsworth, CA). After recovery of the DNA, Klenow enzyme is used to trim single-stranded, nucleotide overhangs creating blunt ends which facilitate religation 15 and cloning.

After ethanol precipitation, the products are redissolved in 13  $\mu$ l of ligation buffer, 1  $\mu$ l T4-DNA ligase (15 units) and 1  $\mu$ l T4 polynucleotide kinase are added, and the mixture is 20 incubated at room temperature for 2-3 hours or overnight at 16°C. Competent *E. coli* cells (in 40  $\mu$ l of appropriate media) are transformed with 3  $\mu$ l of ligation mixture and cultured in 80  $\mu$ l of SOC medium (Sambrook et al., supra). After incubation for one hour at 37°C, the whole 25 transformation mixture is plated on Luria Bertani (LB)-agar (Sambrook et al., supra) containing 2x Carb. The following day, several colonies are randomly picked from each plate and cultured in 150  $\mu$ l of liquid LB/2x Carb medium placed in an individual well of an appropriate, commercially-available, sterile 96-well microtiter plate. The following day, 5  $\mu$ l of each 30 overnight culture is transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5  $\mu$ l of each sample is transferred into a PCR array.

For PCR amplification, 18  $\mu$ l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction are added to each well. Amplification is performed using the following 35 conditions:

Step 1	94°C for 60 sec
Step 2	94°C for 20 sec
Step 3	55°C for 30 sec
Step 4	72°C for 90 sec
35 Step 5	Repeat steps 2-4 for an additional 29 cycles
Step 6	72°C for 180 sec
Step 7	4°C (and holding)

Aliquots of the PCR reactions are run on agarose gels together with molecular weight markers. The sizes of the PCR products are compared to the original partial cDNAs, and appropriate clones are selected, ligated into plasmid, and sequenced.

## VI Labeling and Use of Hybridization Probes

5 Hybridization probes derived from SEQ ID NO:2 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base-pairs, is specifically described, essentially the same procedure is used with larger cDNA fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences), labeled by combining 50 pmol of each oligomer and 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P]

10 adenosine triphosphate (Amersham) and T4 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are substantially purified with Sephadex G-25 superfine resin column (Pharmacia & Upjohn). A portion containing  $10^7$  counts per minute of each of the sense and antisense oligonucleotides is used in a typical membrane based hybridization analysis of human genomic DNA digested with one of the following endonucleases (Ase I, Bgl II, Eco RI,

15 Pst I, Xba I, or Pvu II; DuPont NEN®).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5%  
20 sodium dodecyl sulfate. After XOMAT ART™ film (Kodak, Rochester, NY) is exposed to the blots in a Phosphoimager cassette (Molecular Dynamics, Sunnyvale, CA) for several hours, hybridization patterns are compared visually.

## VII Antisense Molecules

Antisense molecules to the NANCH-encoding sequence, or any part thereof, is used to  
25 inhibit *in vivo* or *in vitro* expression of naturally occurring NANCH. Although use of antisense oligonucleotides, comprising about 20 base-pairs, is specifically described, essentially the same procedure is used with larger cDNA fragments. An oligonucleotide based on the coding sequences of NANCH, as shown in Figures 1A, 1B and 1C, is used to inhibit expression of naturally occurring NANCH. The complementary oligonucleotide is designed from the most  
30 unique 5' sequence as shown in Figures 1A, 1B and 1C and used either to inhibit transcription by preventing promoter binding to the upstream nontranslated sequence or translation of an NANCH-encoding transcript by preventing the ribosome from binding. Using an appropriate

portion of the signal and 5' sequence of SEQ ID NO:2, an effective antisense oligonucleotide includes any 15-20 nucleotides spanning the region which translates into the signal or 5' coding sequence of the polypeptide as shown in Figures 1A, 1B and 1C.

### VIII Expression of NANCH

5 Expression of NANCH is accomplished by subcloning the cDNAs into appropriate vectors and transforming the vectors into host cells. In this case, the cloning vector, pSport, previously used for the generation of the cDNA library is used to express NANCH in *E. coli*. Upstream of the cloning site, this vector contains a promoter for  $\beta$ -galactosidase, followed by sequence containing the amino-terminal Met, and the subsequent seven residues of  
10  $\beta$ -galactosidase. Immediately following these eight residues is a bacteriophage promoter useful for transcription and a linker containing a number of unique restriction sites.

Induction of an isolated, transformed bacterial strain with IPTG using standard methods produces a fusion protein which consists of the first eight residues of  $\beta$ -galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of  
15 NANCH into the bacterial growth media which can be used directly in the following assay for activity.

### IX Demonstration of NANCH Activity

NANCH can be expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with an eukaryotic expression vector encoding NANCH. Eukaryotic expression vectors  
20 are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A small amount of a second plasmid, which expresses any one of a number of marker genes such as  $\beta$ -galactosidase, is co-transformed into the cells in order to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to  
25 allow expression and accumulation of NANCH and  $\beta$ -galactosidase.

Transformed cells expressing  $\beta$ -galactosidase are stained blue when a suitable colorimetric substrate is added to the culture media under conditions that are well known in the art. Stained cells are tested for differences in membrane conductance due to chloride ions or other anions by electrophysiological techniques that are well known in the art. Untransformed  
30 cells, and/or cells transformed with either vector sequences alone or  $\beta$ -galactosidase sequences alone, are used as controls and tested in parallel.

Cells expressing NANCH will have higher anionic conductance than control cells. The

contribution of NANCH to the anionic conductance can be confirmed by incubating the cells using antibodies specific for NANCH. The NANCH-specific antibodies will bind to the extracellular side of NANCH and thereby block the pore in the ion channel.

#### X Production of NANCH Specific Antibodies

5 NANCH that is substantially purified using PAGE electrophoresis (Sambrook, *supra*), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. The amino acid sequence deduced from SEQ ID NO:2 is analyzed using DNASTAR software (DNASTAR Inc) to determine regions of high immunogenicity and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to 10 those of skill in the art. Selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions, is described by Ausubel et al. (*supra*), and others.

Typically, the oligopeptides are 15 residues in length, synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry, and coupled to keyhole limpet hemocyanin (KLH, Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-15 hydroxysuccinimide ester (MBS; Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radioiodinated, goat anti-rabbit IgG.

#### XI Purification of Naturally Occurring NANCH Using Specific Antibodies

20 Naturally occurring or recombinant NANCH is substantially purified by immunoaffinity chromatography using antibodies specific for NANCH. An immunoaffinity column is constructed by covalently coupling NANCH antibody to an activated chromatographic resin, such as CnBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

25 Media containing NANCH is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of NANCH (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/NANCH binding (e.g., a buffer of pH 2-3 or a high concentration of a chaotrope, such as urea or thiocyanate ion), and NANCH is collected.

#### 30 XII Identification of Molecules Which Interact with NANCH

NANCH or biologically active fragments thereof are labeled with <sup>125</sup>I Bolton-Hunter reagent (Bolton et al. (1973) *Biochem. J.* 133: 529). Candidate molecules previously arrayed in

the wells of a multi-well plate are incubated with the labeled NANCH, washed and any wells with labeled NANCH complex are assayed. Data obtained using different concentrations of NANCH are used to calculate values for the number, affinity, and association of NANCH with the candidate molecules.

5 All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited  
10 to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION

(i) APPLICANT: INCYTE PHARMACEUTICALS, INC.

(ii) TITLE OF THE INVENTION: NOVEL HUMAN ANION CHANNEL

(iii) NUMBER OF SEQUENCES: 4

## (iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
- (B) STREET: 3174 Porter Drive
- (C) CITY: Palo Alto
- (D) STATE: CA
- (E) COUNTRY: USA
- (F) ZIP: 94304

## (v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Diskette
- (B) COMPUTER: IBM Compatible
- (C) OPERATING SYSTEM: DOS
- (D) SOFTWARE: FastSEQ for Windows Version 2.0

## (vi) CURRENT APPLICATION DATA:

- (A) PCT APPLICATION NUMBER: To Be Assigned
- (B) FILING DATE: Herewith
- (C) CLASSIFICATION:

## (vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/792,014
- (B) FILING DATE: 31-JAN-1997

## (viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Billings, Lucy J.
- (B) REGISTRATION NUMBER: 36,749
- (C) REFERENCE/DOCKET NUMBER: PF-0206 PCT

## (ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 650-855-0555
- (B) TELEFAX: 650-845-4166
- (C) TELEX:

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 253 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: Consensus
- (B) CLONE: Consensus

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Ala Leu Ser Met Pro Leu Asn Gly Leu Lys Glu Glu Asp Lys Glu  
 1 5 10 15  
 Pro Leu Ile Glu Leu Phe Val Lys Ala Gly Ser Asp Gly Glu Ser Ile  
 20 25 30  
 Gly Asn Cys Pro Phe Ser Gln Arg Leu Phe Met Ile Leu Trp Leu Lys  
 35 40 45  
 Gly Val Val Phe Ser Val Thr Thr Val Asp Leu Lys Arg Lys Pro Ala  
 50 55 60  
 Asp Leu Gln Asn Leu Ala Pro Gly Thr His Pro Pro Phe Ile Thr Phe  
 65 70 75 80  
 Asn Ser Glu Val Lys Thr Asp Val Asn Lys Ile Glu Glu Phe Leu Glu  
 85 90 95  
 Glu Val Leu Cys Pro Pro Lys Tyr Leu Lys Leu Ser Pro Lys His Pro  
 100 105 110  
 Glu Ser Asn Thr Ala Gly Met Asp Ile Phe Ala Lys Phe Ser Ala Tyr  
 115 120 125  
 Ile Lys Asn Ser Arg Pro Glu Ala Asn Glu Ala Leu Glu Arg Gly Leu  
 130 135 140  
 Leu Lys Thr Leu Gln Lys Leu Asp Glu Tyr Leu Asn Ser Pro Leu Pro  
 145 150 155 160  
 Asp Glu Ile Asp Glu Asn Ser Met Glu Asp Ile Lys Phe Ser Thr Arg  
 165 170 175  
 Lys Phe Leu Asp Gly Asn Glu Met Thr Leu Ala Asp Cys Asn Leu Leu  
 180 185 190  
 Pro Lys Leu His Ile Val Lys Val Val Ala Lys Lys Tyr Arg Asn Phe  
 195 200 205  
 Asp Ile Ser Lys Glu Met Thr Gly Ile Trp Arg Tyr Leu Thr Asn Ala  
 210 215 220  
 Tyr Ser Arg Asp Gly Phe Thr Asn Thr Cys Pro Ser Asp Lys Glu Val  
 225 230 235 240  
 Glu Ile Ala Tyr Ser Asp Val Ala Lys Arg Leu Thr Lys  
 245 250

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 819 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: Consensus
- (B) CLONE: Consensus

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GCAGCCGAGC	CGGCCATGGC	GTTGTCGATG	CCGCTGAATG	GGCTGAAGGA	GGAGGACAAA	60
GAGCCCCCTCA	TCGAGCTCTT	CGTCAAGGCT	GGCAGTGATG	GTGAAAGCAT	AGGAAACTGCG	120
CCCTTTTCCC	AGAGGCTCTT	CATGATTCTT	TGGCTCAAAG	GAGTTGTATT	TAGTGTGACG	180
ACTGTTGACC	TGAAAAGGAA	GCCAGCAGAC	CTGCAGAACT	TGGCTCCCGG	GACCCACCCA	240
CCATTATATAA	CTTTCAACAG	TGAAGTCAAA	ACGGATGTAA	ATAAGATTGA	GGAATTCTTT	300
GAAGAAGTCT	TATGCCCTCC	CAAGTACTTA	AAGCTTTCAC	CAAACACCC	AGAATCAAAT	360
ACTGCTGGAA	TGGACATCTT	TGCCAAATTTC	TCTGCATATA	TCAAGAATTTC	AAGGCCAGAG	420
GCTAATGAAG	CACTGGAGAG	GGGTCTCCCTG	AAAACCCCTGC	AGAAACTGGA	TGAATATCTG	480
AATTCTCCTC	TCCCTGATGA	AATTGATGAA	AATAGTATGG	AGGACATAAA	GTTTCTACA	540
CGTAAATTTC	TGGATGGCAA	TGAAATGACA	TTAGCTGATT	GCAACCTGCT	GCCCAAACGT	600
CATATTGTCA	AGGTGGTGGC	CAAAAAATAT	CGCAACTTGT	ATATTCAAA	AGAAATGACT	660
GGCATCTGGA	GATACCTAAC	TAATGCCATAC	AGTAGGGACG	GGTCACCAA	TACCTGTCCC	720

AGTGATAAGG AGGTTGAAAT AGCATATAGT GATGTAGCCA AAAGACTCAC CAAGTAAAAT 780  
 CGCGTTGTA AAAGAGGTGT CTTCATGTCT TCCCCTAAG 819

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 437 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank
- (B) CLONE: 289404

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Asn Asp Glu Asn Tyr Ser Thr Thr Ile Tyr Asn Arg Val Gln Thr  
 1 5 10 15  
 Glu Arg Val Tyr Glu Asp Ser Asp Pro Ala Glu Asn Gly Gly Pro Leu  
 20 25 30  
 Tyr Asp Glu Val His Glu Asp Val Arg Arg Glu Asp Asn Leu Tyr Val  
 35 40 45  
 Asn Glu Leu Glu Asn Gln Glu Tyr Asp Ser Val Ala Val Tyr Pro Val  
 50 55 60  
 Gly Arg Gln Gly Arg Thr Ser Ala Ser Leu Gln Pro Glu Thr Gly Glu  
 65 70 75 80  
 Tyr Val Leu Pro Asp Glu Pro Tyr Ser Lys Ala Gln Asp Pro His Pro  
 85 90 95  
 Gly Glu Pro Thr Ala Asp Glu Asp Ile Ser Leu Glu Glu Leu Leu Ser  
 100 105 110  
 Pro Thr Lys Asp His Gln Ser Asp Ser Glu Glu Pro Gln Ala Ser Asp  
 115 120 125  
 Pro Glu Glu Pro Gln Ala Ser Asp Pro Glu Glu Pro Gln Gly Pro Asp  
 130 135 140  
 Pro Glu Glu Pro Gln Glu Asn Gly Asn Glu Met Glu Ala Asp Leu Pro  
 145 150 155 160  
 Ser Pro Ser Ser Phe Thr Ile Gln Asn Ser Arg Ala Phe Ser Thr Arg  
 165 170 175  
 Glu Ile Ser Pro Thr Ser Tyr Ser Ala Asp Asp Val Ser Glu Gly Asn  
 180 185 190  
 Glu Ser Ala Ser Ala Ser Pro Glu Ile Asn Leu Phe Val Lys Ala Gly  
 195 200 205  
 Ile Asp Gly Glu Ser Ile Gly Asn Cys Pro Phe Ser Gln Arg Leu Phe  
 210 215 220  
 Met Ile Leu Trp Leu Lys Gly Val Val Phe Asn Val Thr Thr Val Asp  
 225 230 235 240  
 Leu Lys Arg Lys Pro Ala Asp Leu His Asn Leu Ala Pro Gly Thr His  
 245 250 255  
 Pro Pro Phe Leu Thr Phe Asn Gly Asp Val Lys Thr Asp Val Asn Lys  
 260 265 270  
 Ile Glu Glu Phe Leu Glu Glu Thr Leu Thr Pro Glu Lys Tyr Pro Arg  
 275 280 285  
 Leu Ala Ala Lys His Arg Glu Ser Asn Thr Ala Gly Ile Asp Ile Phe  
 290 295 300  
 Val Lys Phe Ser Ala Tyr Ile Lys Asn Thr Lys Gln Gln Ser Asn Ala  
 305 310 315 320  
 Ala Leu Glu Arg Gly Leu Thr Lys Ala Leu Lys Lys Leu Asp Asp Tyr  
 325 330 335  
 Leu Asn Thr Pro Leu Pro Glu Glu Ile Asp Ala Asp Thr Arg Gly Asp  
 340 345 350

Asp Glu Lys Gly Ser Arg Arg Lys Phe Leu Asp Gly Asp Glu Leu Thr  
 355 360 365  
 Leu Ala Asp Cys Asn Leu Leu Pro Lys Leu His Val Val Lys Ile Val  
 370 375 380  
 Ala Lys Lys Tyr Arg Asn Tyr Asp Phe Pro Ala Glu Met Thr Gly Leu  
 385 390 395 400  
 Trp Arg Tyr Leu Lys Asn Ala Tyr Ala Arg Asp Glu Phe Thr Asn Thr  
 405 410 415  
 Cys Ala Ala Asp Ser Glu Ile Glu Leu Ala Tyr Ala Asp Val Ala Lys  
 420 425 430  
 Arg Leu Ser Arg Ser  
 435

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 210 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank
- (B) CLONE: 895845

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Val Leu Trp Leu Lys Gly Val Thr Phe Asn Val Thr Thr Val Asp  
 1 5 10 15  
 Thr Lys Arg Arg Thr Glu Thr Val Gln Lys Leu Cys Pro Gly Gly Gln  
 20 25 30  
 Leu Pro Phe Leu Leu Tyr Gly Thr Glu Val His Thr Asp Thr Asn Lys  
 35 40 45  
 Ile Glu Glu Phe Leu Glu Ala Val Leu Cys Pro Pro Arg Tyr Pro Lys  
 50 55 60  
 Leu Ala Ala Leu Asn Pro Glu Ser Asn Thr Ala Gly Leu Asp Ile Phe  
 65 70 75 80  
 Ala Lys Phe Ser Ala Tyr Ile Lys Asn Ser Asn Pro Ala Leu Asn Asp  
 85 90 95  
 Asn Leu Glu Lys Gly Leu Leu Lys Ala Leu Lys Val Leu Asp Asn Tyr  
 100 105 110  
 Leu Thr Ser Pro Leu Pro Glu Glu Val Asp Glu Thr Ser Ala Glu Asp  
 115 120 125  
 Glu Gly Val Ser Gln Arg Lys Phe Leu Asp Gly Asn Glu Leu Thr Leu  
 130 135 140  
 Ala Asp Cys Asn Leu Leu Pro Lys Leu His Ile Val Gln Val Val Cys  
 145 150 155 160  
 Lys Lys Tyr Arg Gly Phe Thr Ile Pro Glu Ala Phe Arg Gly Val His  
 165 170 175  
 Arg Tyr Leu Ser Asn Ala Tyr Ala Arg Glu Glu Phe Ala Ser Thr Cys  
 180 185 190  
 Pro Asp Asp Glu Glu Ile Glu Leu Ala Tyr Glu Gln Val Ala Lys Ala  
 195 200 205  
 Leu Lys  
 210

What is claimed is:

1. A substantially purified human anion channel protein comprising the amino acid sequence of SEQ ID NO:1 or fragments thereof.
2. An isolated and purified polynucleotide sequence encoding the human anion channel protein of claim 1.
3. A polynucleotide sequence which hybridizes under stringent conditions to the polynucleotide sequence of claim 2.
4. A hybridization probe comprising the polynucleotide sequence of claim 2.
5. An isolated and purified polynucleotide sequence comprising SEQ ID NO:2 or variants thereof.
6. A polynucleotide sequence which is complementary to claim 2 or variants thereof.
7. A hybridization probe comprising the polynucleotide sequence of claim 6.
8. An expression vector containing the polynucleotide sequence of claim 2.
9. A host cell containing the vector of claim 8.
10. A method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1 the method comprising the steps of:
  - a) culturing the host cell of claim 9 under conditions suitable for the expression of the polypeptide; and
  - b) recovering the polypeptide from the host cell culture.
11. A pharmaceutical composition comprising a substantially purified human anion channel protein comprising the amino acid sequence of SEQ ID NO:1 in conjunction with a suitable pharmaceutical carrier.
12. A purified antibody which binds specifically to the polypeptide of claim 1.
13. A purified agonist which specifically binds to and modulates the activity of the polypeptide of claim 1.
- 25 14. A purified antagonist which specifically binds to and modulates the activity of the polypeptide of claim 1.
15. A pharmaceutical composition comprising the antagonist of claim 14 in conjunction with a suitable pharmaceutical carrier.
- 30 16. A method for treating cancer comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 15.

17. A method for detecting a polynucleotide which encodes human anion channel protein in a biological sample comprising the steps of:

- a) hybridizing the polynucleotide of claim 6 to nucleic acid material of a biological sample, thereby forming a hybridization complex; and
- 5 b) detecting said hybridization complex, wherein the presence of said complex correlates with the presence of a polynucleotide encoding human anion channel protein in said biological sample.

1/7

9	GCA	GCC	GAG	CCG	GCC	ATG	GCG	TTG	TCG	ATG	CCG	CTG	AAT	GGG	CTG	AAG	GAG	GAG
	M	A	L	S	M	P	L	N	G	L	N	G	L	K	E	E	E	
18																		54
63			72					81			90			99				108
	D	K	E	P	L	I	E	L	F	V	K	A	G	S	D	G	E	S
117			126					135			144			153				162
	I	G	N	C	P	F	S	Q	R	L	F	M	I	L	W	L	K	G
171			180					189			198			207				216
	V	V	F	S	V	T	T	V	D	L	K	R	K	P	A	D	L	Q
225			234					243			252			261				270
	N	L	A	P	G	T	H	P	P	F	I	T	F	N	S	E	V	K
279			288					297			306			315				324
	T	D	V	N	K	I	E	E	F	L	E	E	V	L	C	P	P	K
333			342					351			360			369				378
	Y	L	K	L	S	P	K	H	P	E	S	N	T	A	G	M	D	I

FIGURE 1A

387	396	405	414	423	432												
TTT	GCC	AAA	TTC	TCT	GCA												
F	A	K	F	S	A												
TAT	ATC	AAG	AAT	TCA	AGG												
L	E	R	G	L	K												
CTG	GAG	AGG	GGT	CTG	AAA	ACC	CTG	CAG	AAA	CTG	GAT	GAA	TAT	CTG	AAT	GAA	GCA
P	L	P	D	E	I	D	E	N	S	M	E	D	I	K	F	S	T
441	450	459	468	477	486												
495	504	513	522	531	540												
549	558	567	576	585	594												
603	612	621	630	639	648												
657	666	675	684	693	702												
711	720	729	738	747	756												
TTC	ACC	AAT	ACC	TGT	CCC	AGT	GAT	AAG	GAG	GTT	GAA	ATA	GCA	TAT	AGT	GAT	GTA
F	T	N	T	C	P	S	D	K	E	V	E	I	A	Y	S	D	V

FIGURE 1B

765      774      783      792      801      810  
GCC AAA AGA CTC ACC AAG TAA AAT CGC GTT TGT AAA AGA GGT GTC TTC ATG TCT  
A      K      R      L      T      K  
TCC CCT AAG  
819

FIGURE 1C

4 / 7

FIGURE 2A

17	- - - - -	- - - - -	- - - - -	PLIEL FV KAG SD	259592
181	T S Y S A D D V S E G N E S A S A S P E I N L F V K A G T D				GI 289404
2	- - - - -	- - - - -	- - - - -	- - - - -	GI 895845
29	G E S I G N C P F S Q R L F M I L W L K G V V F S V T T V D	259592			
211	G E S I G N C P F S Q R L F M I L W L K G V V F N V T T V D	259592			GI 289404
2	- - - - -	- - - - -	- - - - -	- - - - -	GI 895845
59	L K R K P A D L Q N L A P G T H P P F I L T F N S E V K T D V	259592			
241	L K R K P A D L H N L A P G T H P P F L T F N G D V K T D V	259592			GI 289404
17	T K R R T E T V Q K L C P G G Q L P F L L Y G T E V H T D T	259592			GI 895845
89	N K I E E F L E E V L C P P K Y L K L S P K H P E S N T A G	259592			
271	N K I E E F L E E T L T P E K Y P R L A A K H R E S N T A G	259592			GI 289404
47	N K I E E F L E A V L C P P R Y P K L A A L N P E S N T A G	259592			GI 895845
119	M D I F A K F S A Y I K N S R P E A N E A L E R G L L K T L	259592			
301	I D I F V K F S A Y I K N T K Q Q S N A A L E R G L T K A L	259592			GI 289404
77	L D I F A K F S A Y I K N S N P A L N D N L E K G L L K A L	259592			GI 895845

FIGURE 2B

149	Q K L D E Y L N S P L P D E I D E N S M -	E D I K F S T R K	259592
331	K K L D D Y L N T P L P E E I D A D T R G D D E K G S R R K	GI 289404	
107	K V L D N Y L T S P L P E E V D E T S - A E D E G V S Q R K	GI 895845	
178	F L D G N E M T L A D C N L L P K L H I V K V V A K K Y R N	259592	
361	F L D G D E L T L A D C N L L P K L H V V K I V A K K Y R N	GI 289404	
136	F L D G N E L T L A D C N L L P K L H I V Q V V C K K Y R G	GI 895845	
208	F D I S K E M T G I W R Y L T N A Y S R D G F T N T C P S D	259592	
391	Y D F P A E M T G L W R Y L K N A Y A R D E F T N T C A A D	GI 289404	
166	F T I P E A F R G V H R Y L S N A Y A R E E F A S T C P D D	GI 895845	
238	K E V E I A Y S D V A K R L T K	259592	
421	S E I E L A Y A D V A K R L S R S	GI 289404	
196	S E I E L A Y E Q V A K A L K	GI 895845	

FIGURE 2C

7/7

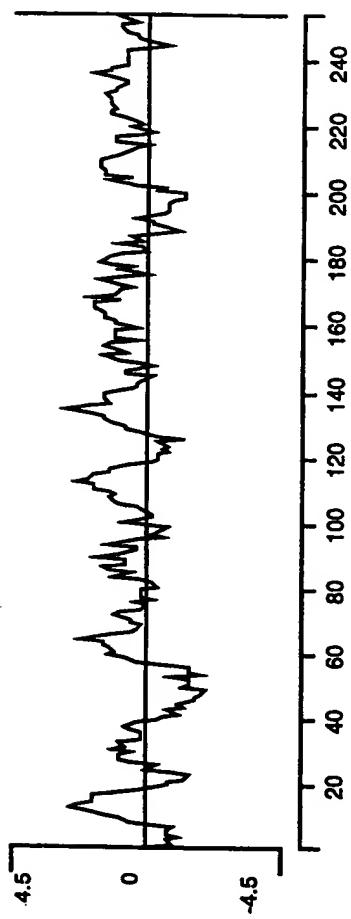


FIGURE 3A

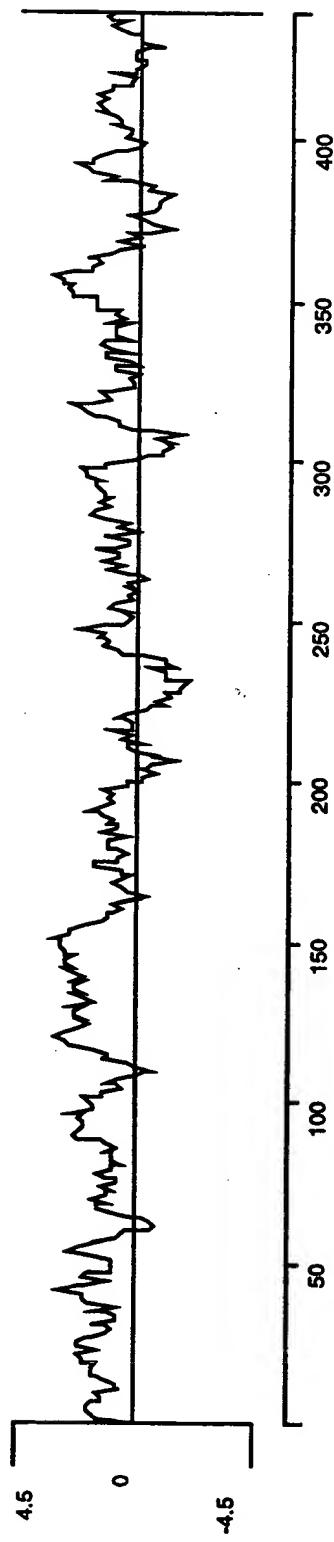


FIGURE 3B

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 98/01422

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 6 C07K14/705

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>D. LANDRY ET AL.: "Molecular Cloning and Characterization of p64, a Chloride Channel Protein from Kidney Microsomes" J.BIOL.CHEM., vol. 268, no. 20, 17 July 1993, USA, pages 14948-55, XP002062699 see page 48, right-hand column - page 49, left-hand column see figure 3</p> <p>-----</p>	3,5-7, 13-15

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed